Cross-frequency temporal envelope correlation for the processing of degraded speech: Preliminary investigations.

Olivier Crouzet, William A. Ainsworth

Cross-frequency Temporal Envelope Correlation in Complex Auditory Signals: Preliminary Investigations.

Olivier Crouzet[1], William A. Ainsworth[2]

[1] *acoustique, acquisition, interprétation (JE2220), Université de Nantes, France*

[2] *Human & Machine Perception Research Centre, MacKay Institute for Communication and Neuroscience, Keele University, UK*

Introduction

When concurrent acoustic signals are produced from different sources, their evolution in the spectro-temporal domain is independent. Each signal may however contain short-term spectral events with a common modulation of amplitude across time. This *co*modulation may be helpful for building auditory streams. When the information that is processed by two spectral channels share a common amplitude modulation, the auditory system may group these events together into a single auditory object whereas when their modulations differ, they would be parsed into different representations.

There is evidence that human listeners are able to perceive the correlation of amplitude modulated signals between spectral channels. When 2 to 5 sine-wave modulated narrow-band signals that differ in centre frequency are produced with phase synchrony in one interval and without synchrony in the other one, listeners can discriminate between the intervals [1, 2]. This result provides evidence that the auditory system is able to process monaural envelope correlation, at least with a small number of sine-wave modulated signals in a discrimination task.

However, little is known about the role of cross-channel correlation for processing complex auditory signals in multi-source environments. If the availability of envelope correlation can help processing such auditory scenes, it is crucial to investigate the actual availability of envelope correlation in natural signals. The present investigation aims at identifying amplitude envelope correlation between spectral bands in natural speech signals by investigating their availability in various amplitude modulation channels.

Method

The aim of this analysis was to investigate the availability of amplitude envelope correlation between spectral channels in natural speech signals. Amplitude modulation was computed for 5 different channels: 0 – 4 Hz, 4 – 8 Hz, 8 – 16 Hz, 16 – 32 Hz and 32 – 64 Hz in each spectral band. Due to the obvious relationship between amplitude envelope and syllabic rhythm, it is expected that the correlation between spectral channels should be stronger for the lowest amplitude modulations (0 - 4 Hz) than for higher ones (from 4 – 8 Hz to 32 – 64 Hz). It is also expected that correlation coefficients should be stronger for close spectral bands.

Table 1: Centre frequency (and Bandwidth, in Hz) of each filterbank channel (the numbers on the left are the channel identifiers used in Tables 2 to 6).

[1]	132 (112)	[5]	1605 (814)
[2]	277 (177)	[6]	2590 (1155)
[3]	548 (366)	[7]	4102 (1871)
[4]	965 (467)	[8]	6427 (2779)

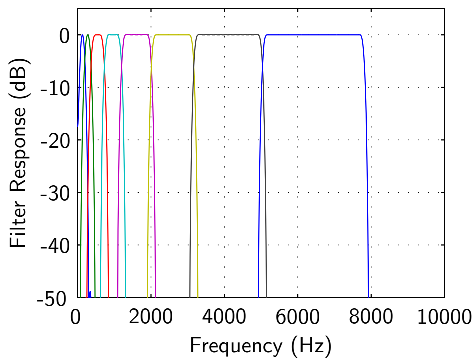

Figure 1: Graphical representation of the 8–channel filterbank response

Material

Thirty 4-digit sequences uttered by male speakers were randomly selected from the TI-Digits database [3]. This database is made of continuous digit sequences pronounced by several American-English speakers in a quiet environment. The TI-Digits sound files are digitized at 20 kHz (16 bit quantization).

Signal processing and analysis

Stimuli were first down-sampled by a factor of 64. They were then passed through an 8-channel FIR filterbank of approximately 1 octave bandwidth with few overlap between filters (cf. Table 1 and Fig. 1). Amplitude envelope was then extracted within each frequency channel by means of a Hilbert transform and half-wave rectified to remove any negative values.

Amplitude modulation channels were then selected by low-pass (0 – 4 Hz) or band-pass filtering (4 – 8 Hz, 8 – 16 Hz, 16 – 32 Hz, 32 – 64 Hz) of the resulting envelope.

Estimates of the correlation between the amplitude of envelope signals were performed by computing the Pearson product-moment correlation (r) [4] between pairs of spectral bands within each modulation channel. Series of one-tailed t-tests were then performed to check whether each of these coefficients were significantly greater than .50.

Results

Results are depicted in Tables 2 to 6. Correlation coefficients that are significantly greater than .50 are printed between square brackets. Non-significant coefficients appear in gray. Within the 0 – 4 Hz interval, strong amplitude envelope correlations are observed between all channel pairs. Each of the computed coefficients is significantly higher than .50 ($p < .05$), which confirms the hypothesis that amplitude envelope correlation is available in natural speech signals.

The observed correlation patterns differ for the remaining amplitude modulation channels (from 4 – 8 Hz to 32 – 64 Hz). Several of the observed coefficients do not reach the significance threshold. However, most if not all coefficients along the diagonal prove to be significantly higher than .50. Therefore, though not all channel pairs exhibit a tendency to comodulate in amplitude when higher amplitude modulations are taken into account, pairs of adjacent channels are still highly correlated to one another.

Table 2: Mean amplitude envelope correlation observed between spectral bands. 0 – 4Hz modulation channel.

	2	3	4	5	6	7	8
1	[0.985]	[0.916]	[0.825]	[0.848]	[0.838]	[0.735]	[0.977]
2		[0.960]	[0.871]	[0.875]	[0.850]	[0.719]	[0.955]
3			[0.960]	[0.918]	[0.860]	[0.688]	[0.899]
4				[0.943]	[0.830]	[0.630]	[0.821]
5					[0.921]	[0.729]	[0.842]
6						[0.904]	[0.868]
7							[0.818]

Table 3: Mean amplitude envelope correlation. 4 – 8Hz modulation channel.

	2	3	4	5	6	7	8
1	[0.933]	[0.748]	[0.559]	0.512	0.529	0.339	[0.918]
2		[0.901]	[0.716]	[0.612]	[0.607]	0.374	[0.831]
3			[0.906]	[0.716]	[0.660]	0.364	[0.687]
4				[0.816]	[0.655]	0.323	0.537
5					[0.820]	0.396	0.497
6						[0.666]	0.566
7							0.450

Table 4: Mean amplitude envelope correlation. 8 – 16Hz modulation channel.

	2	3	4	5	6	7	8
1	[0.860]	[0.617]	0.481	0.425	0.388	0.296	[0.883]
2		[0.875]	[0.703]	0.553	0.506	0.390	[0.730]
3			[0.898]	[0.624]	0.559	0.401	[0.559]
4				[0.723]	0.553	0.355	0.486
5					[0.769]	0.465	0.442
6						[0.761]	0.491
7							0.419

Table 5: Mean amplitude envelope correlation. 16 – 32Hz modulation channel.

	2	3	4	5	6	7	8
1	[0.798]	0.509	0.394	0.310	0.332	0.304	[0.904]
2		[0.861]	[0.720]	0.476	0.462	0.404	[0.701]
3			[0.919]	0.542	0.494	0.406	0.469
4				[0.641]	0.514	0.379	0.393
5					[0.762]	0.496	0.302
6						[0.790]	0.411
7							0.423

Table 6: Mean amplitude envelope correlation. 32 – 64Hz modulation channel.

	2	3	4	5	6	7	8
1	[0.776]	0.479	0.360	0.248	0.283	0.249	[0.849]
2		[0.837]	[0.654]	0.410	0.397	0.338	[0.591]
3			[0.885]	0.508	0.423	0.332	0.365
4				[0.649]	0.467	0.318	0.315
5					[0.711]	0.403	0.216
6						[0.779]	0.397
7							0.466

Discussion

As argued in the introduction, the availability of monaural amplitude envelope correlation between spectral channels may help organize acoustic events into single auditory objects by providing a means to group together sound components that were produced by a common source. Though this research work does not provide any evidence that human listeners may perform such computing, several previous results may be explained by the ability of human listeners to take amplitude modulation correlation into account for the analysis of complex auditory scenes [5, 6, 7].

Dissociations between the 0 – 4 Hz modulation channel and higher ones should be investigated further. This is possible that the latest were simply due to overlap between spectral bands. However, the 8 channel filterbank was built so to minimize the possibility that the same acoustic information would strongly contribute to two adjacent channels.

Eventually, though the 0 – 4 Hz channels data may seem more promising at first, the remaining data look very interesting as well for the understanding of auditory scene analysis.

Acknowledgements

This work was supported by the European Community (SPHEAR, *SPeech HEAring and Recognition* TMR Network).

References

[1] J. Hall III and J. Grose, "Monaural envelope correlation perception in listeners with normal hearing and cochlear impairment," *Journal of Speech & Hearing Research*, vol. 36, 1993.

[2] V. Richards, "Monaural envelope correlation perception," *The Journal of the Acoustical Society of America*, vol. 82, pp. 1621–1630, 1987.

[3] R. Leonard, "A database for speaker-independent digit recognition," in *Proceedings of the ICASSP*, vol. 3, p. 42.11, 1984.

[4] S. van de Par and A. Kohlrausch, "Analytical expressions for the envelope correlation of certain narrow-band stimuli," *The Journal of the Acoustical Society of America*, vol. 98, pp. 3157–3169, 1995.

[5] T. Arai and S. Greenberg, "Speech intelligibility in the presence of cross-channel spectral asynchrony," in *Proceedings of the ICASSP*, (Seattle, USA), 1998.

[6] S. Greenberg and T. Arai, "Speech intelligibility is highly tolerant of cross-channel spectral asynchrony," in *Proceedings of the Joint meeting of the ASA and the ICA*, pp. 2677–2678, 1998.

[7] O. Crouzet and W. Ainsworth, "Envelope information in speech processing: Acoustic-phonetic analysis vs. auditory figure-ground segregation," in *Proceedings of Eurospeech'2001*, (Aalborg, Denmark), 2001.